



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12N 15/00, 15/10, 15/12</b> <b>A61K 35/20</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 92/21752</b> <b>(43) International Publication Date:</b> 10 December 1992 (10.12.92)
<b>(21) International Application Number:</b> PCT/US92/04012 <b>(22) International Filing Date:</b> 21 May 1992 (21.05.92)  <b>(30) Priority data:</b> 707,502                      31 May 1991 (31.05.91)                      US  <b>(71) Applicant:</b> THE UNITED STATES OF AMERICA represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institutes of Health, Office of Technology Transfer, Patent Branch, Bethesda, MD 20892 (US).  <b>(72) Inventors:</b> TENG, Christina ; 8709 Lakewood Drive, Raleigh, NC 27613 (US). PANELLA, Timothy, J. ; 6827 Lindal Road, Knoxville, TN 37931 (US).		<b>(74) Agents:</b> SCOTT, Watson, T. et al.; Cushman, Darby & Cushman, Ninth Floor, 1100 New York, N.W., Washington, DC 20005-3918 (US).  <b>(81) Designated States:</b> AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> HUMAN LACTOFERRIN  <b>(57) Abstract</b>  The present invention relates to a human lactoferrin cDNA gene obtained from human breast tissue and the protein encoded therefrom. The present invention further relates to methods for detecting malignancy arising from tissues that normally secrete lactoferrin using the cDNA gene probe of the present invention. Another aspect of the present invention relates to the promotor region that regulates the human lactoferrin gene.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	MI	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark	MG	Madagascar		
ES	Spain				

## HUMAN LACTOFERRIN

BACKGROUND OF THE INVENTIONFIELD OF THE INVENTION

The present invention relates to a human lactoferrin gene isolated from breast tissue and to the protein product encoded therein. The present invention further relates to the promotor region of human lactoferrin gene and to methods for detecting and analyzing malignancies arising from tissues that normally secrete lactoferrin using a novel human lactoferrin cDNA gene sequence.

BACKGROUND INFORMATION

Lactoferrin is a single polypeptide molecule (M, 76,000) with sites where two oligosaccharide chains can attach (B.F. Anderson et al., *J. Mol. Biol.* 209:711-734 (1989)). This protein shares significant homology with transferrin, however, its role in iron transport is limited since it binds iron 260 times stronger than transferrin (B.F. Anderson et al., (1989)). Two and possibly three isoforms of lactoferrin have been isolated using an affinity chromatography (P. Furnamski et al., *J. Exp. Med.* 170:415-429 (1989); A. Kijlstra et al., *Current Eye Res.*, 8:581-588 (1989)). Lactoferrin has been shown to inhibit bacterial growth by chelating iron and directly attacking the cell wall (R.T. Ellison et al., *Infect Immun.*, 56:2774-2781 (1988)), contribute to the anemia of chronic disease (Birgens. *Scand. J. Haematol.*, 33:225-230 (1984)), improve intestinal absorption of iron in infants (Birgens., (1984)) inhibit myelopoiesis (H.E. Broxmeyer et al., *Blood Cells*

13:31-48 (1987)), and degrade mRNA (P. Furmanski et al., (1989); M.R. Das et al., *Nature* 262:802-805 (1976); P. Furmanski and Z.P. Li, *Exp. Hematol* 18:932-935 (1990). Large quantities of lactoferrin are  
5 found in breast milk (B. Lonnerdal et al., *Nutrition Report Int*, 13:125-134 (1976)), in estrogen-stimulated uterine epithelium (B.T. Pentecost and C.T. Teng, *J. Biol. Chem.* 262:10134-10139 (1987)), and in neutrophilic granulocytes (P.L. Masson et al., *J. Exp. Med.*, 130:643-  
10 658 (1969)) with smaller amounts in tears, saliva, serum, and seminal fluid (D.Y. Mason and C.R. Taylor, *J. Clin. Path.*, 31:316-327 (1978)).

While normal breast ductal epithelium and neutrophilic granulocytes contain lactoferrin, their  
15 malignant counterparts frequently do not (C. Charpin et al., *Cancer*, 55:2612-2617 (1985); T.A. Rado et al., *Blood*, 70:989-993 (1987)). This has been evaluated at the protein level and in a few samples at the messenger RNA level (T.A. Rado et al., (1987)).  
20 Analysis at the genomic level has not been performed. DNA variations, that are detected in the coding regions, may lead to abnormal protein structure and loss of normal function. Variations, such as mutations, deletions, or changes in  
25 methylation, at the promoter regions could lead to altered regulation of the gene. Evaluation of the lactoferrin gene may provide interesting insight concerning the production of lactoferrin in malignant cells. Thus, the need exists for the  
30 structure of the lactoferrin gene including the cDNA and the promotor region. The present invention provides such a description of the structure of a

human lactoferrin cDNA and promotor region of the gene.

Using a lactoferrin cDNA clone isolated from human breast tissue, the applicants have  
5 evaluated restriction fragment length changes in DNA from the white blood cells of 10 normal controls, acute non-lymphocyte leukemia (ANLL) cells from 7 patients, T-cell acute lymphocyte leukemia (ALL) from one patient, 3 leukemia cell lines, and 7  
10 breast cancer cell lines. A comparative study of the lactoferrin gene in these different cell types is provided herein.

The present invention further relates, in part, to a human lactoferrin cDNA and the protein  
15 product encoded therein. In another aspect, the present invention relates to methods for detecting malignancy in tissues that normally secrete lactoferrin by evaluating restriction patterns in DNA using a lactoferrin gene probe of the present  
20 invention.

#### SUMMARY OF THE INVENTION

It is an object of the present invention to provide a DNA sequence of the human lactoferrin gene including the cDNA and the promotor region and  
25 to the protein product encoded therein.

In one embodiment, the present invention relates to a DNA segment encoding human lactoferrin according to the sequence identification number In another embodiment, the present invention relates to  
30 the human lactoferrin protein encoded by the sequences given in identification number 2.

In yet another embodiment, the present invention relates to a DNA segment of the promotor region for human lactoferrin according to the sequence identification number 5 and allelic variations thereof.

In a further embodiment, the present invention relates to a recombinant DNA construct comprising the DNA segments encoding the human lactoferrin gene sequences described above and a vector.

In another embodiment, the present invention relates to a recombinant DNA construct comprising the DNA segment encoding the human lactoferrin gene described above and a DNA promotor regulatory region for human lactoferrin according to sequence identification number 5 or portion thereof operatively linked to the DNA fragment.

In a further embodiment, the present invention relates to a host cell comprising the above described constructs.

Another embodiment of the present invention relates to a method of treating a condition in a patient characterized by a deficiency in lactoferrin by administering to the patient an amount of human lactoferrin according to the present invention in sufficient quantities to eliminate the deficiency. The conditions include neutropenia, AIDS, skin infection, gastrointestinal bacterial overgrowth syndrome, vaginal infection and septic shock.

In yet another embodiment, the present invention relates to methods of diagnosing malignancy or detecting the recovery of a malignancy

from a biological sample comprising the steps of isolating DNA from the biological sample and from normal control samples, cutting the DNA with a restriction enzyme called Xba I, hybridizing the cut DNA with a DNA segment of the human lactoferrin gene of the present invention described above or portion thereof under conditions such that hybridization is effected and comparing the hybridization product patterns of the biological sample and the normal control sample with each other.

In a further embodiment, the present invention relates to a method for detecting small insertions, deletions or mutations surrounding the human lactoferrin gene comprising the steps of isolating the DNA from a biological sample suspected of having such an insertion, deletion or mutation, amplifying the DNA using the human lactoferrin gene segment of the present invention described above or portion thereof in a polymerase chain reaction followed by enzymatically cutting the amplified DNA with Xba I, and hybridizing this DNA with the human lactoferrin gene segment described above under conditions such that hybridization is effected and sequencing the hybridized DNA.

Various other objects and advantages of the present invention will become obvious from the drawings and detailed description of the invention.

The entire contents of all publications mentioned herein are hereby incorporated by reference.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the immunocytochemical staining of normal bone marrow (A) x 400, and breast cancer cell line SKB R3 (B) x 680 using anti-lactoferrin antibody at 1:1500.

- 5     Figure 2 depicts the restriction fragments produced with DNA from normal cells (A) or from leukemia cells (B) using lactoferrin cDNA (HLF 1212) as the probe. Normal samples (n=9) and DNA from 10 different leukemia cells types were digested with indicated enzyme, run in one gel and representative lanes cut out for comparison.

- 15     Figure 3 depicts the restriction fragments produced using DNA from normal samples (A) and from breast cancer cell lines (B), using lactoferrin cDNA (HLF 1212) as a probe. Normal samples (n=2) and DNA from eight cancer lines were digested with indicated enzyme, run in the same gel, and representative lanes cut out for comparison.

- 20     Figure 4 shows the restriction fragments produced using Msp I and lactoferrin cDNA (HLF 1212) as the probe. Lanes 1 - 9 are DNA from normal donors. Lanes 10 - 16 represent DNA from leukemia cells from patients. Lane 17 is cell line K562, lane 18 is KG 1, and lane 19 is U937.

- 25     Figure 5 represents the restriction fragments produced using Msp I and lactoferrin cDNA (HLF 1212) as the probe. Lanes 1 and 2 are DNA from normal donors. Lanes 3 - 9 represent DNA from breast



cancer cell lines. The cell lines are in the following order: Lane 3 - MDAMB 468, lane 4 - MCF 7, lane 5 - BT 474, lane 6 - HBL 100, lane 7 - MDA 175, lane 8 - SKB R3, lane 9 - ZR 75-1.

5 Figure 6 shows the restriction fragments produced using Xba I and lactoferrin cDNA (HLF 1212) as the probe. Lanes 1 - 9 are DNA from normal donors. Lanes 10 - 16 are DNA from leukemia cells from patients and lanes 17 - 19 DNA from leukemia cell  
10 lines (lane 17 - K562, lane 18 - KG1, lane 19 - U937). Arrow A is the band found in patterns A (lanes 1, 2, and 7), B, and C. Arrow B is the band found in patterns B (lanes 3 - 6, 8 - 10, 13, 14) and C. Arrow C is only found in pattern C (lanes 11,  
15 12, 16). Insert is the same specimens run on a 0.7% agarose gel.

Figure 7 depicts the restriction fragments produced using Xba I and lactoferrin cDNA (HLF 1212) as the probe. Lanes 1 and 2 are DNA from normal donors.  
20 Lanes 3 - 9 are DNA from breast cancer cell lines. The order is: Lane 3 - MDAMB 468, lane 4 - BT 474, lane 5 - HBL 100, lane 6 - MDA 175, lane 7 - SKB R3, lane 8 - ZR 75-1, lane 9 - ZR 75-30. Restriction fragment patterns as discussed in the text are in  
25 the following lanes: pattern A is seen in lane 1, pattern B in lane 2, and pattern D in lanes 3 - 9.

Figure 8 shows the restriction fragments produced using Hpa II and lactoferrin cDNA (HLF 1212) as the probe. Lanes 1 - 9 are DNA from normal donors.  
30 Lanes 10 - 16 are DNA from leukemia cells from

patients. Lane 17 is cell line KG1, lane 18 is U937, and lane 19 is HL 60.

Figure 9 shows the restriction fragments produced using Hpa II and lactoferrin cDNA (HLF 1212) as the probe. Lanes 1 and 2 are DNA from normal donors. Lanes 3 - 10 are breast cancer cell lines in the following order: lane 3 - MDAMB 468, lane 4 - MCF 7, lane 5 - BT 474, lane 6 - HBL 100, lane 7 - MDA 175, lane 8 - SKB R3, lane 9 - ZR 75-1, lane 10 - ZR 75-30.

Figure 10 depicts a sequence data of HLF 1212. Differences between the published protein derived AA sequence and our cDNA derived sequence are indicated by underlining the extra AA in our sequence or indicating substitutions beneath our sequence. Nucleotide differences based on published sequence data are indicated above our sequence. Nucleotide changes resulting in a different AA are typed below the area of substitution.

## 20 DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a cDNA sequence for human lactoferrin and the protein encoded therein. The cDNA called HLF1212 was isolated from human breast tissue and is 2117 kb in length. The sequence agrees with the modified amino acid sequence of iron-binding lactoferrin in all areas except the 3 sites in the N-terminal region. One further change is in arginine in place of a lysine at amino acid 200.

30 Another aspect of the present invention relates to methods for diagnosing malignancy by

restriction fragment length polymorphisim (RFLP) analysis of DNA extracted from normal peripheral blood and leukemia cells from patients using the cDNA of the present invention as the probe.

5 Southern analysis indicates that the human lactoferrin gene is polymorphic when tested using Msp I and Xba I restriction enzymes. Further analysis indicates that the changes in the XbaI recognition site could be explained by alterations  
10 in DNA caused by or resulting in malignancy. In the present invention, the DNA from normal and malignant cells are digested with XbaI and the fragment pattern compared using methods well known in the art. The Xba I restriction is associated with 4  
15 patterns in normal and malignant cells (Example 3 and Figures 6 and 7). The most striking change is the deletion of many bands found only in DNA obtained from malignant cells or cell lines derived from either leukemia or breast cancer.

20 If the patterns found in Example 3 (Xba I RFLP pattern C + D) are found in women before breast cancer occurs, it may be easy to screen women at high risk of breast cancer for these changes using cDNA probe of the present invention and RFLP  
25 methodologies well known in the art. For example, lymphocytes may be separated from peripheral blood, DNA extracted, and cut with XbaI. This DNA can then be probed with HLF 1212 or a small piece of HLF 1212 and patterns determined. High risk patients may be  
30 placed on preventive medicines such as Tamoxifen retinoids or have surgery. The same may hold for other hormonally responsive tumors such as prostate, uterus, or tumors arising from

lactoferrin secreting organs such as leukemia, or salivary gland.

Another aspect of the present invention relates to RFLP methods to measure the prognosis of certain types of cancer patients that are given  
5 therapeutics. One may place patients with breast, prostate, uterine, or salivary cancer into risk groups. Those with a specific pattern may be at different risks of disease reoccurrence. Thus, RFLP  
10 analysis using the cDNA probe of the present invention may provide prognostic information for patients with cancer.

Another aspect of the present invention relates to methods for detecting small insertions,  
15 deletions or mutations surrounding the human lactoferrin gene. Either of the above described RFLP methods could be combine with polymerase chain reaction (PCR) analysis. The abnormal area of the gene may be amplified using methods well known in  
20 the art and then mutations detected using restriction analysis (i.e. Xba I) and sequencing.

Yet another aspect of the present invention relates to methods for detecting tumors in pathological specimens that may contain too few  
25 malignant cells to be detected by standard methods. This method may involve PCR of DNA extracted from specimens (biopsy of tissue or bone marrow) and subsequent analysis using the RFLP techniques and DNA probes described above and in the Examples.

30 In another embodiment, the present invention relates to the cDNA clone for human lactoferrin called HLF 1213 and the protein encoded therein. The sequence of HLF 1213 (sequence ID

NO:3) is a combination of clones HLF 1212 (sequence ID NO: 1), 031A (sequence ID NO: 5) and other clones isolated in the same method as HLF 1212. (See Example 2). This clone is a composite of the  
5 complete human lactoferrin cDNA. This clone may be constructed by splicing 2 clones together with HLF 1212 (031A, and HLF 1212). Both HLF 1212 or this combined fragment called HLF 1213 may be used to make recombinant human lactoferrin.

10 In another embodiment, the present invention relates to the human lactoferrin protein obtained from HLF 1212 and HLF 1213 called sequence ID Numbers 2 and 4 respectively.

In yet another embodiment, the present  
15 invention relates to recombinant human lactoferrin expressed in vitro through molecular genetic engineering technology.

The present invention also relates to the recombinant DNA molecules and to host cells  
20 transformed therewith. Using standard methodology well known in the art and described briefly below, a recombinant DNA molecule comprising a vector, for example, a Baculovirus transfer vector and a DNA fragment encoding human lactoferrin, for example,  
25 HLF 1212 or 1213, can be constructed without undue experimentation.

The methods of choice is the Baculovirus-insect cell expression system (M.D. Summers and G.E. Smith, *Texas Agriculture Experiment Station Bulletin* No. 1555, (1987);  
30 V.A. Luckow et al., *Bio/technology* 6:47-55 (1988)). This system has been used successfully to produce commercial quantities of recombinant mammalian glycoproteins. Other expression systems known in

the art can also be used to produce the recombinant protein, for example, yeast, bacterial or mammalian cells.

The 2.2 Kb Eco-R1 fragment containing the entire human lactoferrin coding region may be removed from plasmid HLF 1212 or HLF 1213. The lactoferrin cDNA may be subcloned into Baculovirus transfer vector pAc 700 series (T. Maniatis et al., *Molecular Cloning*: a laboratory manual, Cold Spring Harbor Laboratories, Cold Spring Harbor, New York). Recombinant plasmid (Achlf) may be co-transfected into Sf9 cells along with wild-type AcNPV viral DNA by calcium phosphate transfection procedure (M.D. Summus and G.E. Smith). In vivo homologous recombination between the polyhedron sequences in the wild type viral DNA and the recombinant plasmid results in the generation of recombinant viruses coding for a fused gene product. The recombinant viruses may be plaque purified by screening for the occlusion negative (polyhedron negative) phenotype or by colony hybridization using <sup>32</sup>P-DNA probes covering the HLF-coding region. Characterization of the recombinant viral DNA may be carried out as described by Maniatis et al. Sf9 cells may be plated in 24-well dishes (Costar) at  $3 \times 10^5$  cells/well and allowed to attach for 2 hours in complete Grace's medium. Cells are then infected with wild type AcNPV or recombinant virus AchLF. Two days post-infection, the cell layer and the condition medium may be collected and assayed for the presence of hLF. HLF can be analyzed by SDA-PAGE and Western blotting. Iron binding capacity and anti-bacterial activity may also be examined.

The present invention further relates to treatment of antibacterial and antiviral infections using pharmaceutical doses of human lactoferrin of the present invention (HLF 1212 and 1213  
5 corresponding to sequence ID Nos. 2 and 4 respectively) or recombinant human lactoferrin protein of the present invention.

The actions of lactoferrin are varied; the best established function is antibacterial (R.R. Arnold et al., *Science* 197:263-265 (1977)). Patients  
10 have been found whose neutrophils are deficient in lactoferrin (K.J. Lomax et al., *J. Clin. Invest.* 83:514-519 (1989)). These patients are prone to recurrent infections. Lactoferrin also has been found to  
15 decrease release of CSF or monokines, enhancement monocyte natural killer activity, enhancement of hydroxyl radical production and modulate the activation of the complement system (Birgens, *Scand. J. Haematol* 33:225-230 (1984)). There is also early in  
20 vivo evidence of lactoferrin antiviral activity.

In the past few years, HIV infection has become a significant health problem. HIV causes morbidity by crippling the body's defense mechanism and allowing development of opportunistic  
25 infections. Present treatment is less than ideal and involves treating opportunistic infections as they occur or inhibiting reverse transcriptase. Human lactoferrin is the natural product of the human defense machinery and has been given to  
30 patients both orally and intravenously with no side effects. Due to its bacteriocidal, antifungal, and immunoregulatory activity, administering pharmaceutical acceptable doses of lactoferrin of

the present invention could prove an effective agent to treat patients with AIDS or patients with neutropenia.

Other possible uses of the human lactoferrin of the present invention include treatment of lactoferrin in pharmaceutical doses, either orally or intravenously to patients with skin infections (burn patients), gastrointestinal bacterial overgrowth syndromes, vaginal infections, septic shock, and numerous other disorders.

In yet another embodiment, the present invention relates to the genomic human lactoferrin promotor region (sequence ID No: 5). This sequence contains the entire human lactoferrin promotor region fragment including exon 1 of human lactoferrin clone 1212.

The 5' genomic regulatory region of the present invention has the ability to regulate DNA in a tissue specific manner, i.e., it can be on in breast tissue and off in skin. It also can be hormonally regulated, i.e., on in mid-cycle menstrual cycle, off at menses. This regulation ability may be used in several ways. Genes targeted for transgenic mice may use the lactoferrin promotor. Genes to be used in therapy of human disease (gene therapy) may be linked to the lactoferrin promotor and thus the therapeutic gene regulated in a tissue specific or hormonal pattern.

The invention is described in further detail in the following non-limited examples.



EXAMPLES

The following procedures and materials were used throughout the Examples.

Human tissue.

5                   150 ml of heparinized blood or 5 ml  
heparinized bone marrow was obtained from normal  
paid donors after informed consent was obtained.  
Informed consent and leukemia cells were obtained  
from seven patients with acute leukemia undergoing  
10 emergent leukapheresis. The FAB classification of  
the patients were: two patients with M2, two  
patients with M7, and one patient each with M4, M7,  
ANLL not further specified, and T-cell ALL.  
Nucleated cells were obtained from 80 ml of blood  
15 from normal donors after first incubating cells at  
37° C for 30 min. in 1:20 diluted methylcellulose  
(30 g/500 ml Hank balanced salt solution (HBSS) to  
sediment the red blood cells. The leukocyte-rich  
fraction was removed, and centrifuged into a pellet  
20 at 500 x g for 10 min. at 4° C. Cells from patients  
with leukemia were either used fresh or diluted in  
RPMI 1640 containing 20% fetal calf serum and 10%  
dimethylsulfoxide (DMSO), then frozen at -70° C  
until use. Human leukocyte antigen (HLA) typing,  
25 cytogenetic analysis, and bone marrow biopsy results  
were available for all but one patient who died  
shortly after leukapheresis. All cell lines were  
originally obtained from ATCC (Rockville, MD) and  
maintained at 37° C, 93% humidity, and 5% CO<sub>2</sub>.  
30 Breast cancer cell lines and HBL 100 (a cell line  
derived from a lactating breast) were maintained and  
provided by Dr. J. Dirk Iglehart (Department of

Surgery, Duke University). Cells were grown to confluence and separated from dishes with trypsin 0.05%/EDTA (Gibco), washed, and centrifuged. For all samples, DNA was isolated according to standard methodology (W.M. Strauss in Current Protocols in Molecular Biology. F.A. Ausebel, et al., (eds.), pp. 2.2.1 - 2.2.3 1990. Greene Publishing and Wiley-Interscience, New York.

#### Isolation of cDNA

A Clontech cDNA library from normal human breast tissue (HL 1037b) was plated in host cells Y1090, filter-lifted and probed with mouse lactoferrin cDNA T267 (B.T. Pentecost and C.T. Teng, (1987)). Positive clones were plaque-purified, and the inserts subcloned into the Eco R1 site of Bluescript II SK+ (Stratagene). The recombinant clones were transformed into XL1 Blue cells (Stratagene). A 2.1 Kb insert (HLF 1212) was isolated and sequenced using the dideoxy nucleotide termination reaction and [<sup>32</sup>S]dATP label under contract by Lark sequencing company.

#### Southern Analysis

Ten µg of DNA was digested at 37° C for three hours with Eco R1, Bam H1, Hind III, Pvu II, Pst I, Msp I, Xba I, Hpa II, Mbo I or Sau 3AI under conditions specified by the manufacturer (BRL). Hpa II and Sau 3AI will not cleave DNA when specific bases within their recognition sites are methylated. Msp I and Mbo I respectively, recognize these same sites and are methylation insensitive. DNA was loaded into 0.7, 0.8, or 1.2% agarose gels, run

overnight, and transferred either to Genescreen Plus (nylon, DuPont) or BA-S NC (nitrocellulose, Schleicher & Schuel). Lactoferrin cDNA was removed from plasmid with Eco RI, redigested with Pst I, and gel purified. Both fragments were labeled with [<sup>32</sup>P]dCTP using a random primer kit (Stratagene) to a specific activity of 1 x 10<sup>8</sup>. Hybridization was performed exactly according to Genescreen instructions or a modification of BA-S NC instructions (hybridization solution - 50% formamide, 5x SSPE, 1% SDS, 4x Denhardt, 100 µg/ml single stranded DNA, 7.5% dextran, pre-hybridization solution - the same as above with 5% formamide and no dextran). Filters were washed at high stringency at 60° C and exposed to Kodak XOMAT AR film using intensifying screens for 3-7 days. DNA from normal and leukemic cells was probed with histone cDNA (Oncore) as a control; no polymorphic pattern was found.

#### Immunocytochemistry

Antibody against human milk lactoferrin (Sigma) was raised in rabbits and the IgG fraction was prepared as described previously (C.T. Teng et al., *Endocrinology* 124:992-999 (1989)). All cell lines, normal cells, and leukemia patient's cells were examined using this antibody. Ten normal bone marrow specimens were stained to define the specific cell in bone marrow that begins to produce lactoferrin. Cells were smeared onto alcohol-washed, pre-cleaned slides, air dried 1 hour, and fixed in 95% methanol, and 1.7% formalin for 10 min. Slides were next rinsed in dH<sub>2</sub>O and either air dried and stored in a moisture proof container at 4° C or

used immediately. Staining procedure was followed directions provided with Vector ABC-AP kit using levamisol as the blocking agent, antibody dilution of 1:1500, and hematoxylin (gill #3) counterstain.  
5 Three-hundred cells per sample were scored manually as negative, trace, or positive.

Example 1. Immunocytochemical staining.

As shown in Table 1 and Figure 1A, bone marrow lactoferrin began to appear in the myelocyte stage with almost all cells staining positively by  
10 the metamyelocyte stage. None of the leukemia cells from patients or leukemia cell lines contained stainable lactoferrin. Occasional positive granulocytes could be seen in with the leukemic  
15 cells from patients. Breast cancer cell lines stained negatively for lactoferrin except for 1.5% trace positive cells in SKB R3 (Figure 1B).

Table 1. Immunocytochemical staining of normal bone marrow using anti-lactoferrin antibody

	Blasts and Promyelocytes	Myelocytes	Metamyelocytes	Bands	Neutrophils
Negative	93% <sup>a</sup> (8.6)	30% (20.4)	12% (7.5)	3% (1.2)	1% (1)
Trace	6% (8.2)	38% (8.3)	40% (10.6)	10% (5.2)	2% (2)
Positive	0.3% (0.4)	32% (19.2)	48% (17)	88% (4.5)	97% (2)

<sup>a</sup> - values represent the mean of 10 bone marrow samples stained with the standard deviation in parenthesis, >300 cells counted per sample.

Example 2. Library screening, isolation and characterization of HLF 1212 clone.

Thirty human lactoferrin clones were isolated from the breast tissue cDNA library. The longest (HLF 1212) was sequenced completely. This clone is 2117 bp's in length and includes a 17 amino acid (AA) leader sequence (no ATG site) and is 4 AA shy of the 3' terminus (Figure 10). The AA sequence coded for by HLF 1212 has 4 sites that differ from the previously published revised AA sequence derived from the protein (B.F. Anderson et al., (1989)). In the sequence of the present invention, there is one insertion (Arginine (Arg) at AA 22, bp 64-6) and three substitutions (Glutamine (Gln) for Asparagine (Asn) at AA 31, bp 91-3; Isoleucine (Ile) for Leucine (Leu) at AA 55, bp 163-5; and Arg for Lysine (Lys) at AA 218, bp 652-4). The first three of these changes are clustered at the 5' end. Contained within HLF 1212, but not in any of the 10 other partially sequenced isolates, is a deleted cytosine at bp 2097 (AA 699) which caused a frame-shift at the 3' end of the protein. This extra base was confirmed by repeated bi-directional sequencing. The deletion at 2097 is now thought to be either a cloning artifact or a rare species of mRNA.

In addition to cDNA of the present invention, three other authors have published lactoferrin cDNA sequence data (T.A. Rado, et al., (1987); M.J. Powell and J.E. Ogden, *Nucleic Acids Res.*, 18:4013, (1990); M.W. Rey et al., *Nucleic Acids Res.*, 18:5288, (1990)). All of these sequences are different, and a comparison between the AA data derived from the protein and sequence changes derived from the cDNA, are presented in Figure 10. When compared to HLF 1212, all of the sequences

contain an extra cytosine at bp 2097 (AA 699).  
Powell et al., (1990) isolated a 2.3 kb sequence  
from breast tissue that, except for the extra  
cytosine, is identical to our cDNA in the areas of  
5 overlap. The isolate of the present invention  
differs from that of Rado's 3' 1023 base fragment in  
4 locations (T.A. Rado et al., (1987)) with one  
resulting difference in the AA sequence (Gly for Ala  
at AA 486, bp 1456-8). Two silent mutations and the  
10 extra cytosine make up the remainder of the changes.  
Ray et al have also published a cDNA sequence  
isolated from human mammary tissue that contains two  
AA changes (Ile for Thr at AA 147, bp 440-2; and Gly  
for Cys at AA 421, bp 1261-3) and one silent base  
15 difference (M.W. Rey et al., (1990)).

Example 3. Evaluation of restriction fragments  
using lactoferrin HLF 1212 as probe.

The fragments produced by digestion with  
Eco RI, Bam HI, Hind III, Pst I, Pvu II, Sau 3AI, or  
20 Mbo I, were nearly identical whether the DNA was  
from normal or malignant cells. The fragment  
patterns produced by these restriction enzymes in  
DNA from leukemic and breast cancer cells are shown  
in Figures 2 and 3. Restriction with Msp I  
25 indicated the deletion of a 3.5 Kb band in 3 of 10  
leukemic cells (Figure 4), 4 of 7 breast cancer cell  
lines (Figure 5), and a much fainter hybridization  
of this band in 2 of 9 normal specimens (Figure 4).  
An extra 1.3 Kb band also occurred in the breast  
30 cancer line MDA 175 (Figure 5, lane 7). There was  
no relationship between the phenotype or chromosome

analysis of the leukemia patients and the Msp I changes.

Fragments produced by Xba I fell into 4 patterns. All patterns contained 4 unchanged bands (~6.5 kb, ~4.2 kb, ~3.0 kb, and ~2.2 kb). Pattern A occurred in 3 of 9 normal samples and contained a 3.5 Kb band and three light < 2.0 kb bands in addition to the unchanged bands (Figure 6, lanes 1, 2, and 7; Figure 7, lane 1). Pattern B was seen in 6 of 9 normal and 3 of 7 leukemia cells from patients and contained extra 3.5, 5.0, and 6.7 Kb bands along with the three light < 2.0 kb bands and the unchanged bands (Figure 6, lanes 3-6, 8, 9, 10, 13, 14; Figure 7, lane 2). The last patterns were only seen in DNA obtained from malignant tissue. In pattern C, an extra 9.0 Kb band together with the 3.5, 5.0, and 6.6 kb and unchanged bands were observed in three leukemia patient samples (Figure 6 lanes 11, 12 (see insert) and lane 16). Also noted is the absence of the light < 2.0 kb bands. Pattern D contained only the 4 unchanged and the three light < 2.0 kb bands and was present in DNA obtained from all three leukemia and all seven breast cancer cell lines, (Figure 6, lanes 17 - 19, and Figure 7, lanes 3 - 9). There was one patient (M2 leukemia) with a restriction pattern like that of the cell lines (Figure 6, lane 15). There were no chromosomal abnormalities, French-American-British (FAB) categories, or phenotypic types associated with any polymorphic Xba I pattern.



Example 4. Isolation and characterization of the genomic lactoferrin promotor region.

A human placental DNA library (Clontech) was plated on LE 392 bacterial cells and screened and probed with the 5' end of HLF 1212 (1.3Kb). Positive clones were cut with SAC 1 and rescreened using a 25 base oligonucleotide (synthesized to match Exon 1 of p1212). All SAC 1 fragments from clone 031A were transformed into Bluescript II KS (stratagene) plasmid. Clone 031A-30 was 2.0 kb and hybridized to Exon 1 oligonucleotide probe. This was sequenced using dideoxynucleotide chain termination and synthesized oligonucleotide primers. Sequence ID NO. 5 shows the sequence of the entire fragment (5' - 3') that includes Exon 1.

\* \* \* \*

While the foregoing invention has been described in some detail for purpose of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Teng, Christina  
Panella, Timothy J.
- (ii) TITLE OF INVENTION: HUMAN LACTOFERRIN
- (iii) NUMBER OF SEQUENCES: 5
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
  - (B) STREET: 1615 L. STREET N.W., ELEVENTH FLOOR
  - (C) CITY: WASHINGTON
  - (D) STATE: D.C.
  - (E) COUNTRY: USA
  - (F) ZIP: 20036-5601
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: SCOTT, WATSON T.
  - (B) REGISTRATION NUMBER: 26,581
  - (C) REFERENCE/DOCKET NUMBER: WTS/5683/84482/KIK
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (202) 861-3000
  - (B) TELEFAX: (202) 822-0944
  - (C) TELEX: 6714627 CUSH

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2117 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

25

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..2117

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTT	GTC	TTC	CTC	GTC	CTG	CTG	TTC	CTC	GGG	GCC	CTC	GGA	CTG	TGT	CTG	48
Leu	Val	Phe	Leu	Val	Leu	Leu	Phe	Leu	Gly	Ala	Leu	Gly	Leu	Cys	Leu	
1				5					10					15		
GCT	GGC	CGT	AGG	AGA	AGG	AGT	GTT	CAG	TGG	TGC	GCC	GTA	TCC	CAA	CCC	96
Ala	Gly	Arg	Arg	Arg	Arg	Ser	Val	Gln	Trp	Cys	Ala	Val	Ser	Gln	Pro	
			20					25					30			
GAG	GCC	ACA	AAA	TGC	TTC	CAA	TGG	CAA	AGG	AAT	ATG	AGA	AAA	GTG	CGT	144
Glu	Ala	Thr	Lys	Cys	Phe	Gln	Trp	Gln	Arg	Asn	Met	Arg	Lys	Val	Arg	
		35					40					45				
GGC	CCT	CCT	GTC	AGC	TGC	ATA	AAG	AGA	GAC	TCC	CCC	ATC	CAG	TGT	ATC	192
Gly	Pro	Pro	Val	Ser	Cys	Ile	Lys	Arg	Asp	Ser	Pro	Ile	Gln	Cys	Ile	
	50					55					60					
CAG	GCC	ATT	GCG	GAA	AAC	AGG	GCC	GAT	GCT	GTG	ACC	CTT	GAT	GGT	GGT	240
Gln	Ala	Ile	Ala	Glu	Asn	Arg	Ala	Asp	Ala	Val	Thr	Leu	Asp	Gly	Gly	
65					70					75					80	
TTC	ATA	TAC	GAG	GCA	GGC	CTG	GCC	CCC	TAC	AAA	CTG	CGA	CCT	GTA	GCG	288
Phe	Ile	Tyr	Glu	Ala	Gly	Leu	Ala	Pro	Tyr	Lys	Leu	Arg	Pro	Val	Ala	
				85					90					95		
GCG	GAA	GTC	TAC	GGG	ACC	GAA	AGA	CAG	CCA	CGA	ACT	CAC	TAT	TAT	GCC	336
Ala	Glu	Val	Tyr	Gly	Thr	Glu	Arg	Gln	Pro	Arg	Thr	His	Tyr	Tyr	Ala	
			100					105					110			
GTG	GCT	GTG	GTG	AAG	AAG	GGC	GGC	AGC	TTT	CAG	CTG	AAC	GAA	CTG	CAA	384
Val	Ala	Val	Val	Lys	Lys	Gly	Gly	Ser	Phe	Gln	Leu	Asn	Glu	Leu	Gln	
		115				120						125				
GGT	CTG	AAG	TCC	TGC	CAC	ACA	GGC	CTT	CGC	AGG	ACC	GCT	GGA	TGG	AAT	432
Gly	Leu	Lys	Ser	Cys	His	Thr	Gly	Leu	Arg	Arg	Thr	Ala	Gly	Trp	Asn	
	130					135					140					
GTC	CCT	ATA	GGG	ACA	CTT	CGT	CCA	TTC	TTG	AAT	TGG	ACG	GGT	CCA	CCT	480
Val	Pro	Ile	Gly	Thr	Leu	Arg	Pro	Phe	Leu	Asn	Trp	Thr	Gly	Pro	Pro	
145					150					155					160	

26

GAG	CCC	ATT	GAG	GCA	GCT	GTG	GCC	AGG	TTC	TTC	TCA	GCC	AGC	TGT	GTT	528
Glu	Pro	Ile	Glu	Ala	Ala	Val	Ala	Arg	Phe	Phe	Ser	Ala	Ser	Cys	Val	
				165					170					175		
CCC	GGT	GCA	GAT	AAA	GGA	CAG	TTC	CCC	AAC	CTG	TGT	CGC	CTG	TGT	GCG	576
Pro	Gly	Ala	Asp	Lys	Gly	Gln	Phe	Pro	Asn	Leu	Cys	Arg	Leu	Cys	Ala	
			180					185					190			
GGG	ACA	GGG	GAA	AAC	AAA	TGT	GCC	TTC	TCC	TCC	CAG	GAA	CCG	TAC	TTC	624
Gly	Thr	Gly	Glu	Asn	Lys	Cys	Ala	Phe	Ser	Ser	Gln	Glu	Pro	Tyr	Phe	
		195					200					205				
AGC	TAC	TCT	GGT	GCC	TTC	AAG	TGT	CTG	AGA	GAC	GGG	GCT	GGA	GAC	GTG	672
Ser	Tyr	Ser	Gly	Ala	Phe	Lys	Cys	Leu	Arg	Asp	Gly	Ala	Gly	Asp	Val	
	210					215					220					
GCT	TTT	ATC	AGA	GAG	AGC	ACA	GTG	TTT	GAG	GAC	CTG	TCA	GAC	GAG	GCT	720
Ala	Phe	Ile	Arg	Glu	Ser	Thr	Val	Phe	Glu	Asp	Leu	Ser	Asp	Glu	Ala	
225					230					235					240	
GAA	AGG	GAC	GAG	TAT	GAG	TTA	CTC	TGC	CCA	GAC	AAC	ACT	CGG	AAG	CCA	768
Glu	Arg	Asp	Glu	Tyr	Glu	Leu	Leu	Cys	Pro	Asp	Asn	Thr	Arg	Lys	Pro	
				245					250					255		
GTG	GAC	AAG	TTC	AAA	GAC	TGC	CAT	CTG	GCC	CGG	GTC	CCT	TCT	CAT	GCC	816
Val	Asp	Lys	Phe	Lys	Asp	Cys	His	Leu	Ala	Arg	Val	Pro	Ser	His	Ala	
				260				265					270			
GTT	GTG	GCA	CGA	AGT	GTG	AAT	GGC	AAG	GAG	GAT	GCC	ATC	TGG	AAT	CTT	864
Val	Val	Ala	Arg	Ser	Val	Asn	Gly	Lys	Glu	Asp	Ala	Ile	Trp	Asn	Leu	
		275					280					285				
CTC	CGC	CAG	GCA	CAG	GAA	AAG	TTT	GGA	AAG	GAC	AAG	TCA	CCG	AAA	TTC	912
Leu	Arg	Gln	Ala	Gln	Glu	Lys	Phe	Gly	Lys	Asp	Lys	Ser	Pro	Lys	Phe	
	290					295					300					
CAG	CTC	TTT	GGC	TCC	CCT	AGT	GGG	CAG	AAA	GAT	CTG	CTG	TTC	AAG	GAC	960
Gln	Leu	Phe	Gly	Ser	Pro	Ser	Gly	Gln	Lys	Asp	Leu	Leu	Phe	Lys	Asp	
305					310					315					320	
TCT	GCC	ATT	GGG	TTT	TCG	AGG	GTG	CCC	CCG	AGG	ATA	GAT	TCT	GGG	CTG	1008
Ser	Ala	Ile	Gly	Phe	Ser	Arg	Val	Pro	Pro	Arg	Ile	Asp	Ser	Gly	Leu	
				325					330					335		
TAC	CTT	GGC	TCC	GGC	TAC	TTC	ACT	GCC	ATC	CAG	AAC	TTG	AGG	AAA	AGT	1056
Tyr	Leu	Gly	Ser	Gly	Tyr	Phe	Thr	Ala	Ile	Gln	Asn	Leu	Arg	Lys	Ser	
			340					345					350			

GAG Glu	GAG Glu	GAA Glu 355	GTG Val	GCT Ala	GCC Ala	CGG Arg	CGT Arg 360	GCG Ala	CGG Arg	GTC Val	GTG Val	TGG Trp 365	TGT Cys	GCG Ala	GTG Val	1104
GGC Gly 370	GAG Glu	CAG Gln	GAG Glu	CTG Leu	CGC Arg	AAG Lys 375	TGT Cys	AAC Asn	CAG Gln	TGG Trp	AGT Ser 380	GGC Gly	TTG Leu	AGC Ser	GAA Glu	1152
GGC Gly 385	AGC Ser	GTG Val	ACC Thr	TGC Cys	TCC Ser 390	TCG Ser	GCC Ala	TCC Ser	ACC Thr	ACA Thr 395	GAG Glu	GAC Asp	TGC Cys	ATC Ile	GCC Ala 400	1200
CTG Leu	GTG Val	CTG Leu	AAA Lys	GGA Gly 405	GAA Glu	GCT Ala	GAT Asp	GCC Ala	ATG Met 410	AGT Ser	TTG Leu	GAT Asp	GGA Gly	GGA Gly	TAT Tyr 415	1248
GTG Val	TAC Tyr	ACT Thr	GCA Ala 420	GGC Gly	AAA Lys	TGT Cys	GGT Gly	TTG Leu 425	GTG Val	CCT Pro	GTC Val	CTG Leu	GCA Ala 430	GAG Glu	AAC Asn	1296
TAC Tyr	AAA Lys	TCC Ser 435	CAA Gln	CAA Gln	AGC Ser	AGT Ser	GAC Asp 440	CCT Pro	GAT Asp	CCT Pro	AAC Asn	TGT Cys 445	GTG Val	GAT Asp	AGA Arg	1344
CCT Pro 450	GTG Val	GAA Glu	GGA Gly	TAT Tyr	CTT Leu	GCT Ala 455	GTG Val	GCG Ala	GTG Val	GTT Val	AGG Arg 460	AGA Arg	TCA Ser	GAC Asp	ACT Thr	1392
AGC Ser 465	CTT Leu	ACC Thr	TGG Trp	AAC Asn	TCT Ser 470	GTG Val	AAA Lys	GGC Gly	AAG Lys	AAG Lys 475	TCC Ser	TGC Cys	CAC His	ACC Thr	GCC Ala 480	1440
GTG Val	GAC Asp	AGG Arg	ACT Thr	GCA Ala 485	GGC Gly	TGG Trp	AAT Asn	ATC Ile	CCC Pro 490	ATG Met	GGC Gly	CTG Leu	CTC Leu	TTC Phe 495	AAC Asn	1488
CAG Gln	ACG Thr	GGC Gly	TCC Ser 500	TGC Cys	AAA Lys	TTT Phe	GAT Asp	GAA Glu 505	TAT Tyr	TTC Phe	AGT Ser	CAA Gln	AGC Ser 510	TGT Cys	GCC Ala	1536
CCT Pro	GGG Gly	TCT Ser 515	GAC Asp	CCG Pro	AGA Arg	TCT Ser	AAT Asn 520	CTC Leu	TGT Cys	GCT Ala	CTG Leu	TGT Cys 525	ATT Ile	GGC Gly	GAC Asp	1584
GAG Glu 530	CAG Gln	GGT Gly	GAG Glu	AAT Asn	AAG Lys	TGC Cys 535	GTG Val	CCC Pro	AAC Asn	AGC Ser	AAC Asn 540	GAG Glu	AGA Arg	TAC Tyr	TAC Tyr	1632

28

GGC TAC ACT GGG GCT TTC CGG TGC CTG GCT GAG AAT GCT GGA GAC GTT Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp Val 545 550 555 560	1680
GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT AAC Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn Asn 565 570 575	1728
AAT GAG GCA TGG GCT AAG GAT TTG AAG CTG GCA GAC TTT GCG CTG CTG Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu Leu 580 585 590	1776
TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAG GCT AGA AGC TGC CAT Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys His 595 600 605	1824
CTT GCC ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT AAG GTG Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys Val 610 615 620	1872
GAA CGC CTG AAA CAG GTG TTG CTC CAC CAA CAG GCT AAA TTT GGG AGA Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe Gly Arg 625 630 635 640	1920
AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA ACC Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu Thr 645 650 655	1968
AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG GCC AGA CTC CAT Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu His 660 665 670	2016
GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC GCA GGC Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala Gly 675 680 685	2064
ATT ACT AAT CTG AAA AAG TGC TCA ACC TCC CCC TCC TGG AAG CCT GTG Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Ser Trp Lys Pro Val 690 695 700	2112
AAT TC Asn 705	2117

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 705 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu Cys Leu
 1           5           10           15
Ala Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser Gln Pro
          20           25           30
Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys Val Arg
          35           40           45
Gly Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln Cys Ile
          50           55           60
Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp Gly Gly
          65           70           75           80
Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro Val Ala
          85           90           95
Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr Tyr Ala
          100          105          110
Val Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu Leu Gln
          115          120          125
Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly Trp Asn
          130          135          140
Val Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly Pro Pro
          145          150          155          160
Glu Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser Cys Val
          165          170          175
Pro Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys Ala
          180          185          190
Gly Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro Tyr Phe
          195          200          205

```

30

Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp Val  
 210 215 220  
 Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp Glu Ala  
 225 230 235 240  
 Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys Pro  
 245 250 255  
 Val Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His Ala  
 260 265 270  
 Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn Leu  
 275 280 285  
 Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys Phe  
 290 295 300  
 Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe Lys Asp  
 305 310 315 320  
 Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly Leu  
 325 330 335  
 Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys Ser  
 340 345 350  
 Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala Val  
 355 360 365  
 Gly Glu Gln Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu Ser Glu  
 370 375 380  
 Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile Ala  
 385 390 395 400  
 Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Tyr  
 405 410 415  
 Val Tyr Thr Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn  
 420 425 430  
 Tyr Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp Arg  
 435 440 445  
 Pro Val Glu Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser Asp Thr  
 450 455 460  
 Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr Ala  
 465 470 475 480



Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu Phe Asn  
 485 490 495  
 Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys Ala  
 500 505 510  
 Pro Gly Ser Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile Gly Asp  
 515 520 525  
 Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr Tyr  
 530 535 540  
 Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp Val  
 545 550 555 560  
 Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn Asn  
 565 570 575  
 Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu Leu  
 580 585 590  
 Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys His  
 595 600 605  
 Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys Val  
 610 615 620  
 Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe Gly Arg  
 625 630 635 640  
 Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu Thr  
 645 650 655  
 Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu His  
 660 665 670  
 Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala Gly  
 675 680 685  
 Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Ser Trp Lys Pro Val  
 690 695 700  
 Asn  
 705

32

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2124

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG AAA CTT GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG	48
Met Lys Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu	
1 5 10 15	
TGT CTG GCT GGC CGT AGG AGA AGG AGT GTT CAG TGG TGC GCC GTA TCC	96
Cys Leu Ala Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser	
20 25 30	
CAA CCC GAG GCC ACA AAA TGC TTC CAA TGG CAA AGG AAT ATG AGA AAA	144
Gln Pro Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys	
35 40 45	
GTG CGT GGC CCT CCT GTC AGC TGC ATA AAG AGA GAC TCC CCC ATC CAG	192
Val Arg Gly Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln	
50 55 60	
TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC CTT GAT	240
Cys Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp	
65 70 75 80	
GGT GGT TTC ATA TAC GAG GCA GGC CTG GCC CCC TAC AAA CTG CGA CCT	288
Gly Gly Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro	
85 90 95	
GTA GCG GCG GAA GTC TAC GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT	336
Val Ala Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr	
100 105 110	

TAT GCC GTG GCT GTG GTG AAG AAG GGC GGC AGC TTT CAG CTG AAC GAA	384
Tyr Ala Val Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu	
115 120 125	
CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC CTT CGC AGG ACC GCT GGA	432
Leu Gln Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly	
130 135 140	
TGG AAT GTC CCT ATA GGG ACA CTT CGT CCA TTC TTG AAT TGG ACG GGT	480
Trp Asn Val Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly	
145 150 155 160	
CCA CCT GAG CCC ATT GAG GCA GCT GTG GCC AGG TTC TTC TCA GCC AGC	528
Pro Pro Glu Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser	
165 170 175	
TGT GTT CCC GGT GCA GAT AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG	576
Cys Val Pro Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu	
180 185 190	
TGT GCG GGG ACA GGG GAA AAC AAA TGT GCC TTC TCC TCC CAG GAA CCG	624
Cys Ala Gly Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro	
195 200 205	
TAC TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AGA GAC GGG GCT GGA	672
Tyr Phe Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly	
210 215 220	
GAC GTG GCT TTT ATC AGA GAG AGC ACA GTG TTT GAG GAC CTG TCA GAC	720
Asp Val Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp	
225 230 235 240	
GAG GCT GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG	768
Glu Ala Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg	
245 250 255	
AAG CCA GTG GAC AAG TTC AAA GAC TGC CAT CTG GCC CGG GTC CCT TCT	816
Lys Pro Val Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser	
260 265 270	

CAT GCC GTT GTG GCA CGA AGT GTG AAT GGC AAG GAG GAT GCC ATC TGG	864
His Ala Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp	
275 280 285	
AAT CTT CTC CGC CAG GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG	912
Asn Leu Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro	
290 295 300	
AAA TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG CTG TTC	960
Lys Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe	
305 310 315 320	
AAG GAC TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT	1008
Lys Asp Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser	
325 330 335	
GGG CTG TAC CTT GGC TCC GGC TAC TTC ACT GCC ATC CAG AAC TTG AGG	1056
Gly Leu Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg	
340 345 350	
AAA AGT GAG GAG GAA GTG GCT GCC CGG CGT GCG CGG GTC GTG TGG TGT	1104
Lys Ser Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys	
355 360 365	
GCG GTG GGC GAG CAG GAG CTG CGC AAG TGT AAC CAG TGG AGT GGC TTG	1152
Ala Val Gly Glu Gln Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu	
370 375 380	
AGC GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC	1200
Ser Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys	
385 390 395 400	
ATC GCC CTG GTG CTG AAA GGA GAA GCT GAT GCC ATG AGT TTG GAT GGA	1248
Ile Ala Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly	
405 410 415	
GGA TAT GTG TAC ACT GCA GGC AAA TGT GGT TTG GTG CCT GTC CTG GCA	1296
Gly Tyr Val Tyr Thr Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala	
420 425 430	

GAG AAC TAC AAA TCC CAA CAA AGC AGT GAC CCT GAT CCT AAC TGT GTG	1344
Glu Asn Tyr Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val	
435 440 445	
GAT AGA CCT GTG GAA GGA TAT CTT GCT GTG GCG GTG GTT AGG AGA TCA	1392
Asp Arg Pro Val Glu Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser	
450 455 460	
GAC ACT AGC CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC TGC CAC	1440
Asp Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His	
465 470 475 480	
ACC GCC GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG GGC CTG CTC	1488
Thr Ala Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu	
485 490 495	
TTC AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC	1536
Phe Asn Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser	
500 505 510	
TGT GCC CCT GGG TCT GAC CCG AGA TCT AAT CTC TGT GCT CTG TGT ATT	1584
Cys Ala Pro Gly Ser Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile	
515 520 525	
GGC GAC GAG CAG GGT GAG AAT AAG TGC GTG CCC AAC AGC AAC GAG AGA	1632
Gly Asp Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg	
530 535 540	
TAC TAC GGC TAC ACT GGG GCT TTC CGG TGC CTG GCT GAG AAT GCT GGA	1680
Tyr Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly	
545 550 555 560	
GAC GTT GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA	1728
Asp Val Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly	
565 570 575	
AAT AAC AAT GAG GCA TGG GCT AAG GAT TTG AAG CTG GCA GAC TTT GCG	1776
Asn Asn Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala	
580 585 590	

CTG CTG TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAG GCT AGA AGC	1824
Leu Leu Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser 595 600 605	
TGC CAT CTT GCC ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT	1872
Cys His Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp 610 615 620	
AAG GTG GAA CGC CTG AAA CAG GTG TTG CTC CAC CAA CAG GCT AAA TTT	1920
Lys Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe 625 630 635 640	
GGG AGA AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT	1968
Gly Arg Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser 645 650 655	
GAA ACC AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG GCC AGA	2016
Glu Thr Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg 660 665 670	
CTC CAT GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC	2064
Leu His Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val 675 680 685	
GCA GGC ATT ACT AAT CTG AAA AAG TGC TCA ACC TCC CCC CTC CTG GAA	2112
Ala Gly Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu 690 695 700	
GCC TGT GAA TTC	2124
Ala Cys Glu Phe 705	

37

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 708 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Lys Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu
 1             5             10             15
Cys Leu Ala Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser
          20             25             30
Gln Pro Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys
      35             40             45
Val Arg Gly Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln
      50             55             60
Cys Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp
 65             70             75             80
Gly Gly Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro
          85             90             95
Val Ala Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr
      100             105             110
Tyr Ala Val Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu
      115             120             125
Leu Gln Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly
      130             135             140
Trp Asn Val Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly
      145             150             155             160
Pro Pro Glu Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser
          165             170             175
Cys Val Pro Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu
      180             185             190
Cys Ala Gly Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro
      195             200             205

```

38

Tyr Phe Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly  
 210 215 220  
 Asp Val Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp  
 225 230 235 240  
 Glu Ala Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg  
 245 250 255  
 Lys Pro Val Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser  
 260 265 270  
 His Ala Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp  
 275 280 285  
 Asn Leu Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro  
 290 295 300  
 Lys Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe  
 305 310 315 320  
 Lys Asp Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser  
 325 330 335  
 Gly Leu Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg  
 340 345 350  
 Lys Ser Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys  
 355 360 365  
 Ala Val Gly Glu Gln Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu  
 370 375 380  
 Ser Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys  
 385 390 395 400  
 Ile Ala Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly  
 405 410 415  
 Gly Tyr Val Tyr Thr Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala  
 420 425 430  
 Glu Asn Tyr Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val  
 435 440 445  
 Asp Arg Pro Val Glu Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser  
 450 455 460  
 Asp Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His  
 465 470 475 480



Thr Ala Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu  
                             485                            490                            495

Phe Asn Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser  
                             500                            505                            510

Cys Ala Pro Gly Ser Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile  
                             515                            520                            525

Gly Asp Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg  
                             530                            535                            540

Tyr Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly  
                             545                            550                            555                            560

Asp Val Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly  
                             565                            570                            575

Asn Asn Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala  
                             580                            585                            590

Leu Leu Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser  
                             595                            600                            605

Cys His Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp  
                             610                            615                            620

Lys Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe  
                             625                            630                            635                            640

Gly Arg Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser  
                             645                            650                            655

Glu Thr Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg  
                             660                            665                            670

Leu His Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val  
                             675                            680                            685

Ala Gly Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu  
                             690                            695                            700

Ala Cys Glu Phe  
 705

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2086 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGAGGATCAT GGCTCACTGC CACCTTCATC TCCCAGGCTC AAATGGTCCT CCCACTTTAG	60
CCTCCCAAGT AGCTGGGACC ATAGGCATAC ACCACCATGC TGGGCTAATT TTTGTATTTT	120
TTGTAGAGAT GGGGGTTTCC CTATGAAGCC CAGGCTAGTC TTGAACTCCT GGGCTCAAGC	180
GATCCTCCCA TCTTGGCCTC CCAAAGTGCT GGGATTACAG GCATGAGCCA CTGTGCCCTG	240
CCTAGTTACT CTTGGGCTAA GTTCACATCC ATACACACAG GATATTCTTT CTGAGGCCCC	300
CAATGTGTCC CACAGGCACC ATGCTGTATG TGCACTCCC CTAGAGATGG ATGTTTAGTT	360
TGCTTCCAAC TGATTAATGG CATGCAGTGG TGCCTGGAAA CATTTGTACC TGGGGTGCTG	420
TGTGTCATGG GAATGTATTT ACGAGATGTA TTCTTAGAAG CAGTATTCTA GCTTTTGAAT	480
TTTAAAATCT GACATTTATG GCGATTGTTA AAATGAGGTT ACCATTTTCCT ACTGAATACT	540
ATCAACACCA AAAAAGAAGA AGGAGGAGAT GGAGAAAAA AAGACAAAA AAAAAAAGT	600
GGTAGGGCAT CTTAGCCATA GGGCATCTTT CTCATTGGCA AATAAGAACA TGGGAACCAGC	660
CTTGGGTGGT GGCCATTCCC CTCTGAGGTC CCTGTCTGTT TTCTGGGAGC TGTATTGTGG	720
GTCTCAGCAG GGCAGGGAGA TACCCCATGG GCAGCTTGCC TGAGACTCTG GGCAGCCTCT	780
CTTTTCTCTG TCAGCTGTCC CTAGGCTGCT GCTGGGGGTG GTCGGGTCAT CTTTTCAACT	840
CTCAGCTCAC TGCTGAGCCA AGGTGAAAGC AAACCCACCT GCCCTAACTG GCTCCTAGGC	900
ACCTTCAAGG TCATCTGCTG AAGAAGATAG CAGTCTCACA GGTCAAGGCG ATCTTCAAGT	960
AAAGACCCTC TGCTCTGTGT CCTGCCCTCT AGAAGGCACT GAGACCAGAG CTGGGACAGG	1020
GCTCAGGGGG CTGCGACTCC TAGGGGCTTG CAGACCTAGT GGGAGAGAAA GAACATCGCA	1080

GCAGCCAGGC AGAACCAGGA CAGGTGAGGT GCAGGCTGGC TTTCCTCTCG CAGCGCGGTG 1140  
TGGAGTCCTG TCCTGCCTCA GGGCTTTTCG GAGCCTGGAT CCTCAAGGAA CAAGTAGACC 1200  
TGGCCGCGGG GAGTGGGGAG GGAAGGGGTG TCTATTGGGC AACAGGGCGG GGCAAAGCCC 1260  
TGAATAAAGG GGCGCAGGGC AGGCGCAAGT GGCAGAGCCT TCGTTTGCCA AGTCGCCTCC 1320  
AGACCGCAGA CATGAAACTT GTCTTCCTCG TCCTGCTGTT CCTCGGGGCC CTCGGTGAGT 1380  
GCAGGTGCCT GGGGGCGCGA GCCGCCTGAT GGGCGTCTCC TGCGCCCTGT CTGCTAGGCG 1440  
CTTTGGTCCC TGTGTCCGGT TGGCTGGGCG CGGGGTCTCT GCGCCCCGCG GTCCCAGCGC 1500  
CTACAGCCGG GAGGCGGCCC GGACGCGGGG CCAGTCTCTT TCCCACATGG GGAGGAACAG 1560  
GAGCTGGGCT CCTCAAGCCG GATCGGGGCA CGCCTAGCTC TGCTCAGAGC TTCTCAAAAG 1620  
GCCTCCCAGG CCCCTGTCCC TTTGTGTCCC GCCTAAGGAT TTGGTCCCCA TTGTATTGTG 1680  
ACATGCGTTT TACCTGGGAG GAAAGTGAGG CTCAGAGAGG GTGAGCGACT AGCTCAAGGA 1740  
CCCTAGTCCA GATCCTAGCT CCTGCGAGGA CTGTGAGACC CCAGCAAGAC CGAGCCTTTA 1800  
TGAGACTTAG TTTCTTCACT TAAAGAAACG GCCTAACCAT GGGTCCACAG GGTGTGAGG 1860  
AGGAGATGGG GCATTGCGAC ACCTTCCGTG GCAGAGGGTT GTGGAGGGGT GCGGTGCTCC 1920  
TGATGGAACC CTGTGTCAGA GGGTTTGAGA GGGAAATGTC AGCCAAACAG AAGGAAGGAG 1980  
CAGAAGGAAG GAAACAATTG TCAGTTCCAT AACCAAAGTA ATTTCTCGGG TGCTCAGAGG 2040  
GCACTCCCCA GCGCTGCACA TTAGTGACCT AAATGCGTGA GTGCGG 2086

**WHAT IS CLAIMED IS:**

1. A DNA segment encoding human lactoferrin according to sequence I.D. No.: 1.
2. Human lactoferrin protein according to sequence I.D. No.: 2.
3. A DNA promotor region for human lactoferrin according to sequence I.D. No.: 5 and allelic variations thereof.
4. A recombinant DNA construct comprising:
  - i) said DNA segment according to claim 1 and
  - ii) a vector
5. The DNA construct according to claim 4 further comprising the regulating sequence according to sequence I.D. No.: 5 or portion thereof operatively linked to said DNA fragment.
6. The DNA construct according to claim 4 or 5 wherein said vector is pAc 700 series.
7. A host cell comprising said DNA construct according to claim 4 or 5.
8. The cell according to claim 7, wherein said host cell is Sf9 cells.
9. A recombinant lactoferrin protein expressed in the host cell of claim 7.
10. A method of treating a condition in a patient characterization by a deficiency in

lactoferrin, administering to said patient an amount of human lactoferrin according to claims 2 or 9 sufficient to eliminate said deficiency.

- 5           11. The method of claim 10 wherein said condition is neutropenia, AIDS, skin infection, gastrointestinal bacterial overgrowth syndrome, vaginal infection or septic shock.
- 10           12. A method of diagnosing malignancy in a biological sample comprising the steps of:
- i) isolating DNA from said biological sample and normal control sample
  - ii) cutting said DNA with restriction enzyme, Xba I,
  - 15       iii) hybridizing said cut DNA of step (ii) with a DNA segment according to claim 1 or 2 or portion thereof under conditions such that hybridization is effected and
  - 20       iv) comparing the hybridization products of step 3 from said biological sample and normal sample to each other.
- 25           13. A method of detecting recovery of a disease in a patient given a therapeutic comprising the steps of:
- i) isolating DNA from a biological sample of said patient and normal human control sample,
  - ii) cutting said DNA with Xba I,
  - 30       iii) hybridizing said cut DNA of step (ii) with a DNA segment according to claim 1 or

portion thereof under conditions such that hybridization is effected and

- iv) comparing the hybridization products of the biological sample in step 3 to the hybridization products of normal sample in step 3 to determine the relatedness to normal samples.

14. A method for detecting insertions, deletions or mutations surrounding the human lactoferrin gene comprising the steps of

- i) isolating DNA from a biological sample suspected of having said insertion, deletion or mutation,
- ii) amplifying said DNA using the DNA fragment of claim 1 or portion thereof in a polymerase chain reaction,
- iii) cutting said amplified DNA with restriction enzyme Xbu I,
- iv) hybridizing said DNA from step (iii) with the DNA fragment according to claim 1 or portion thereof under conditions such that hybridization is effected and
- v) sequencing said DNA of step (iv).

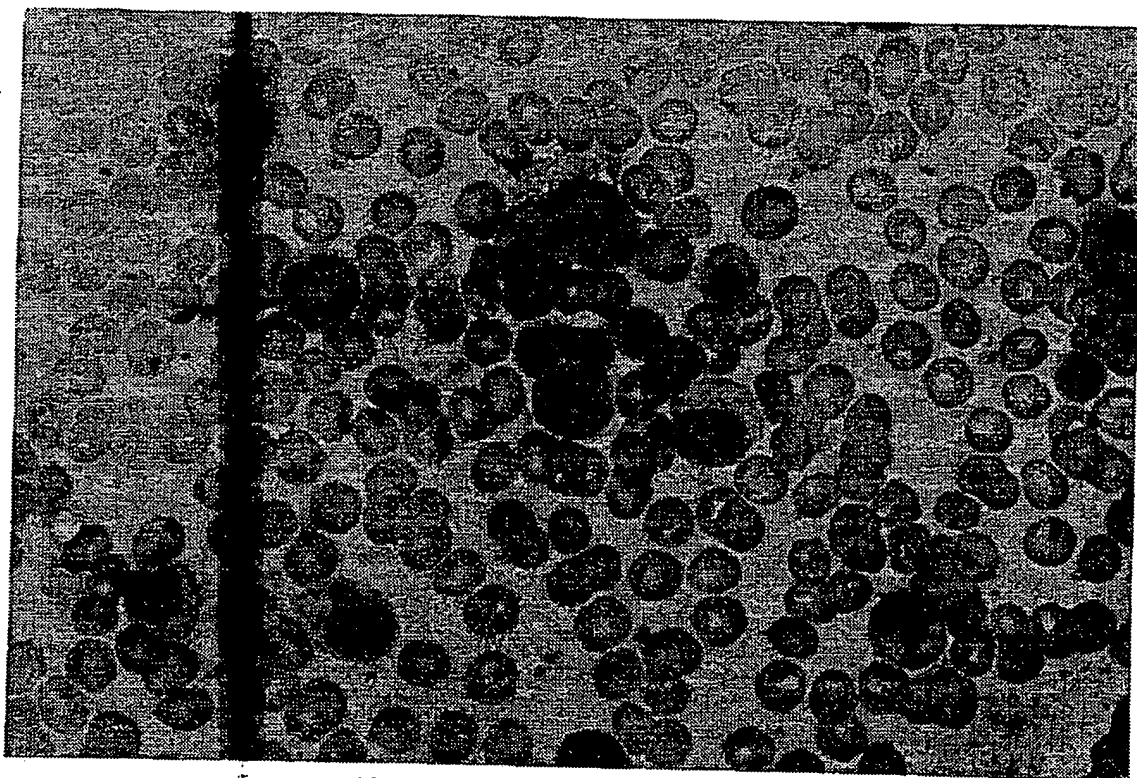


FIG. 1A

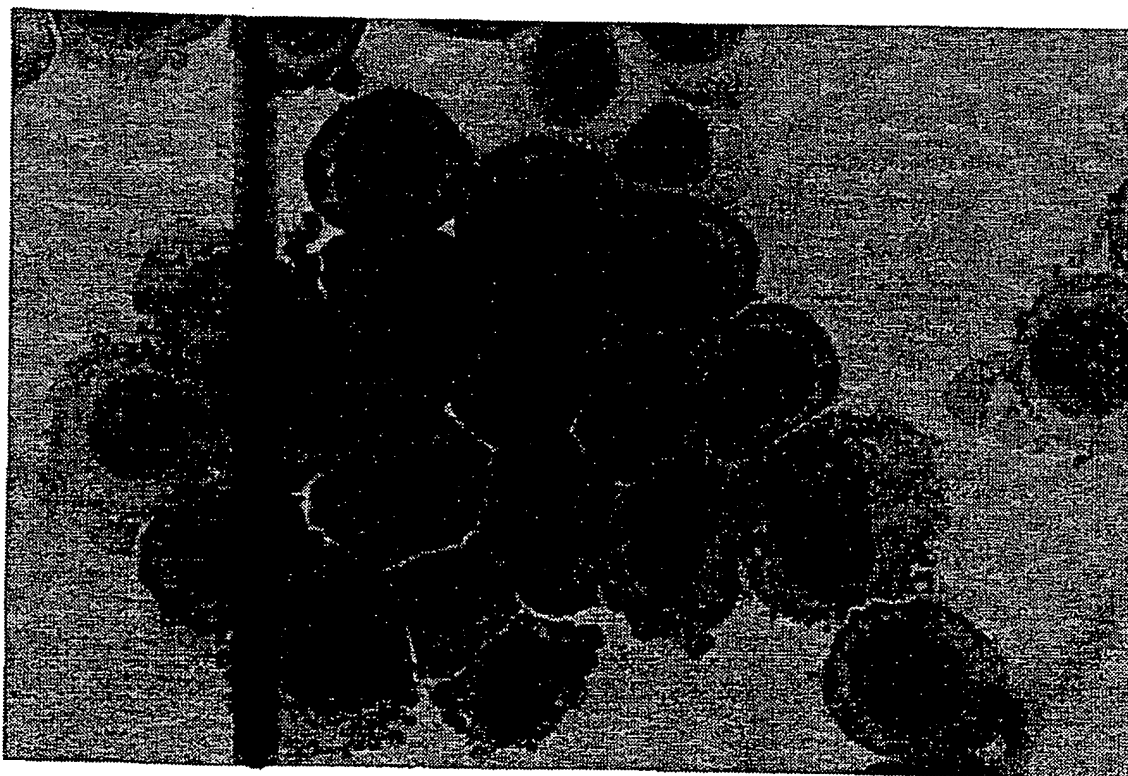


FIG. 1B

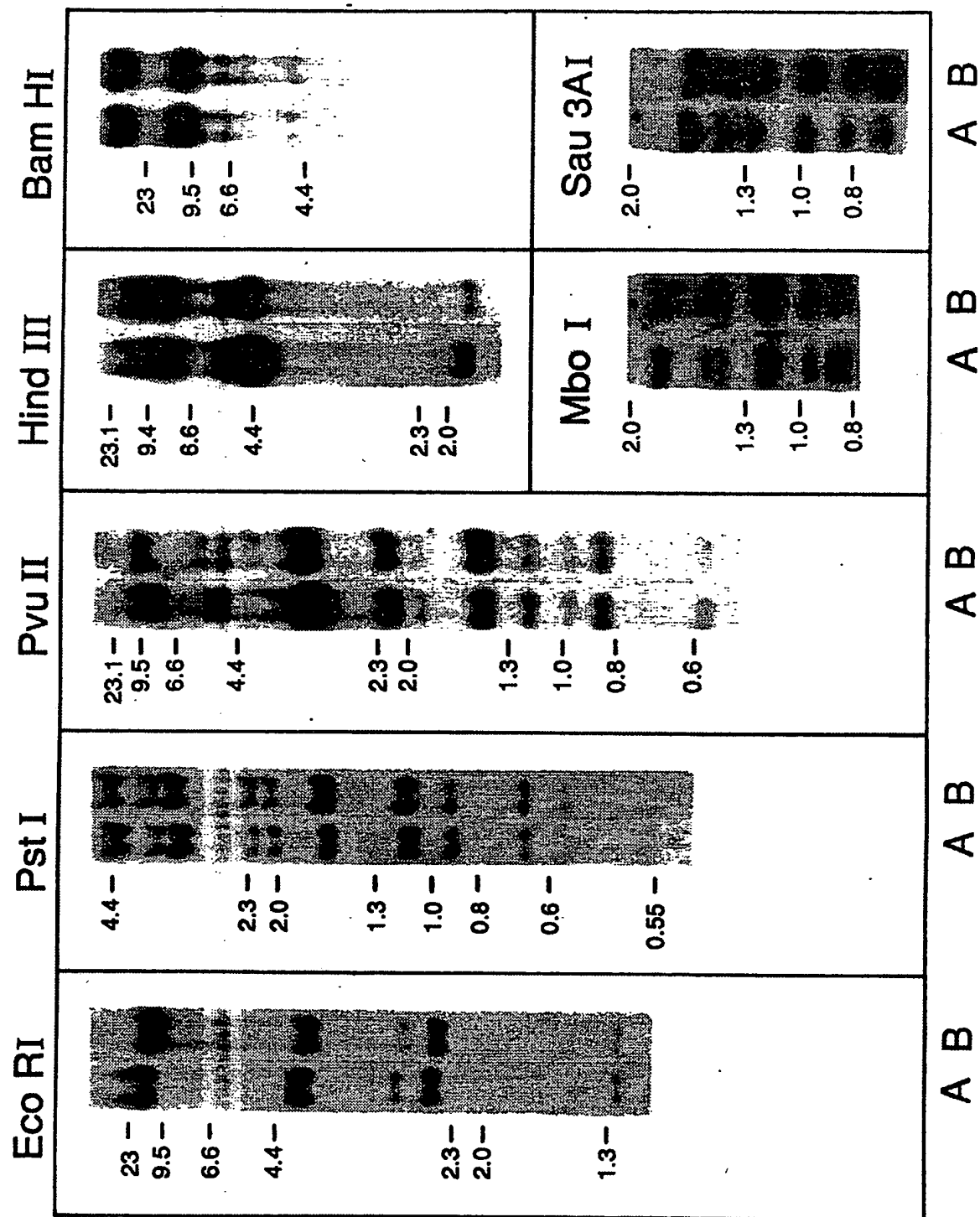


FIG. 2



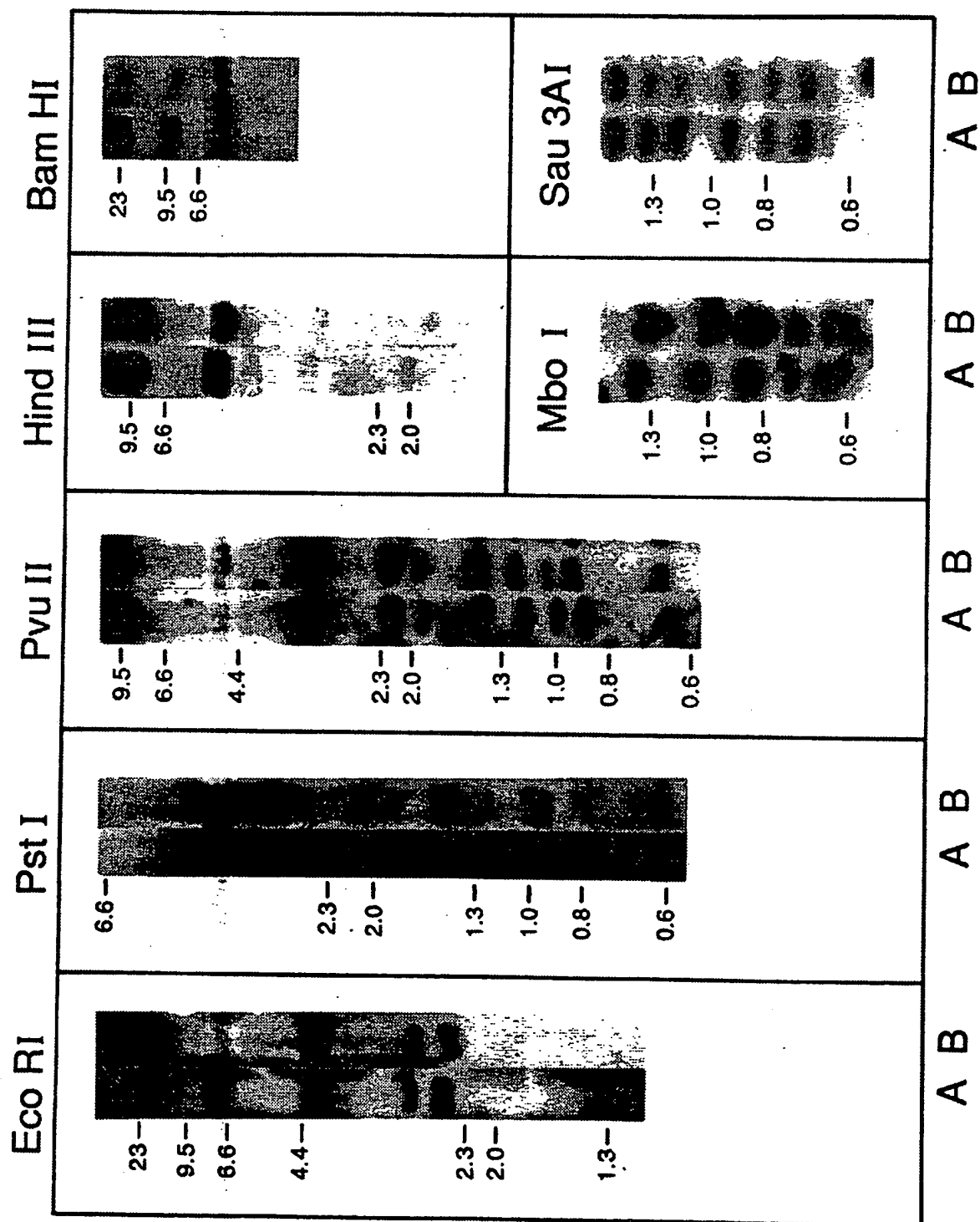


FIG. 3

FIG. 4

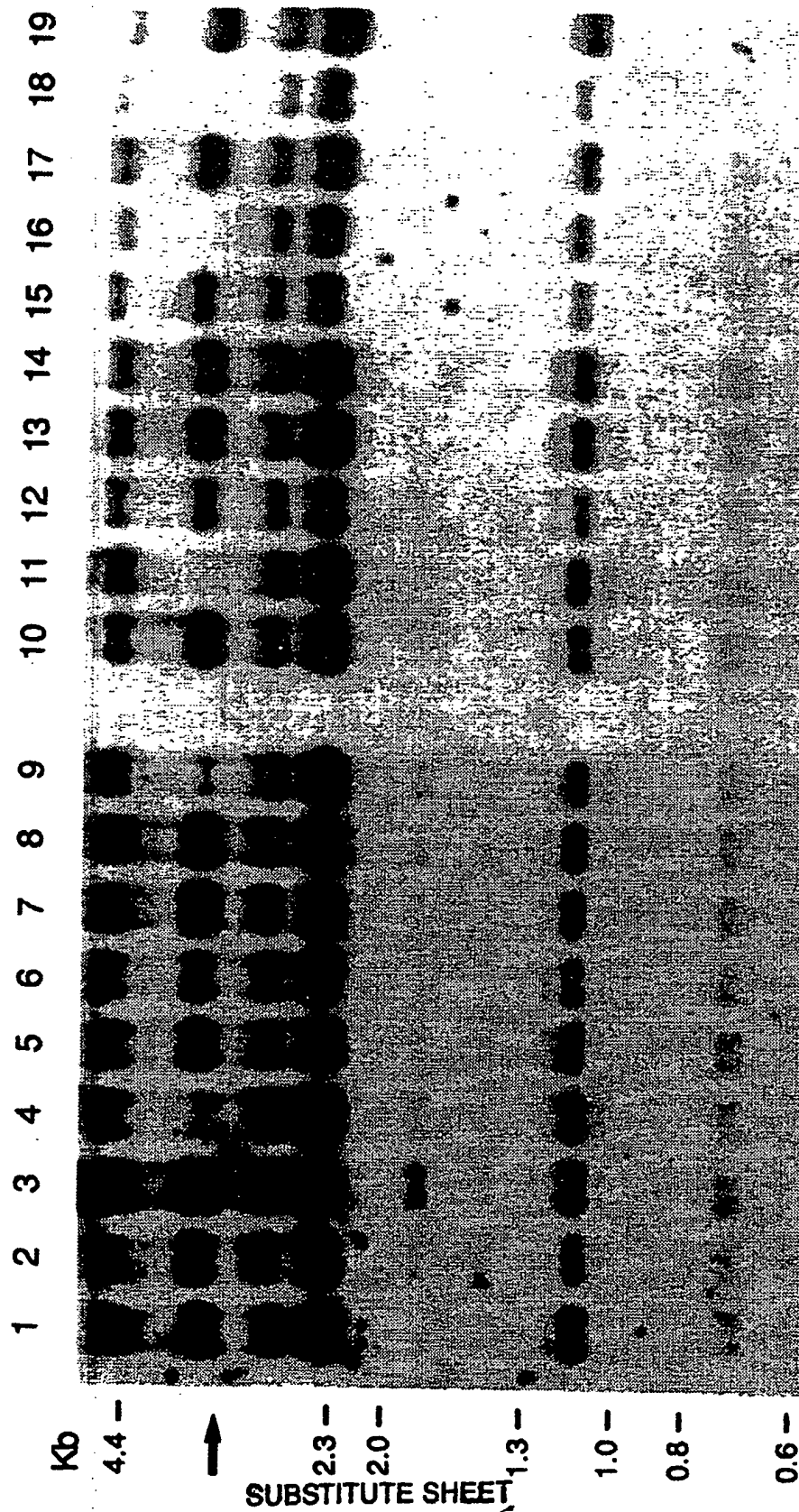
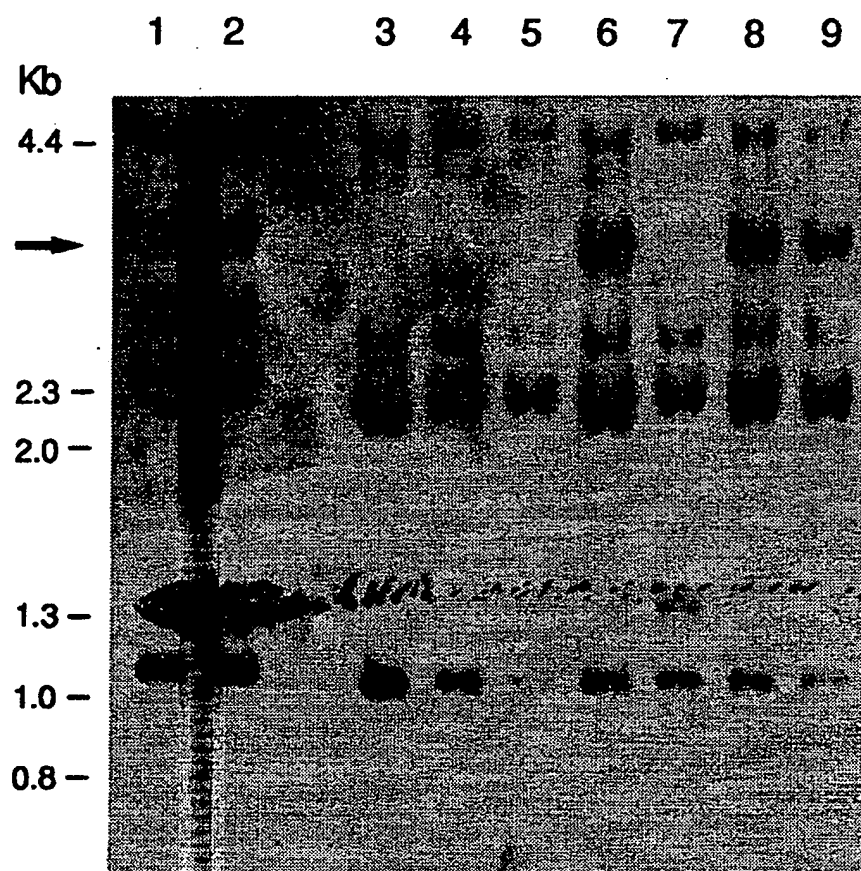


FIG. 5



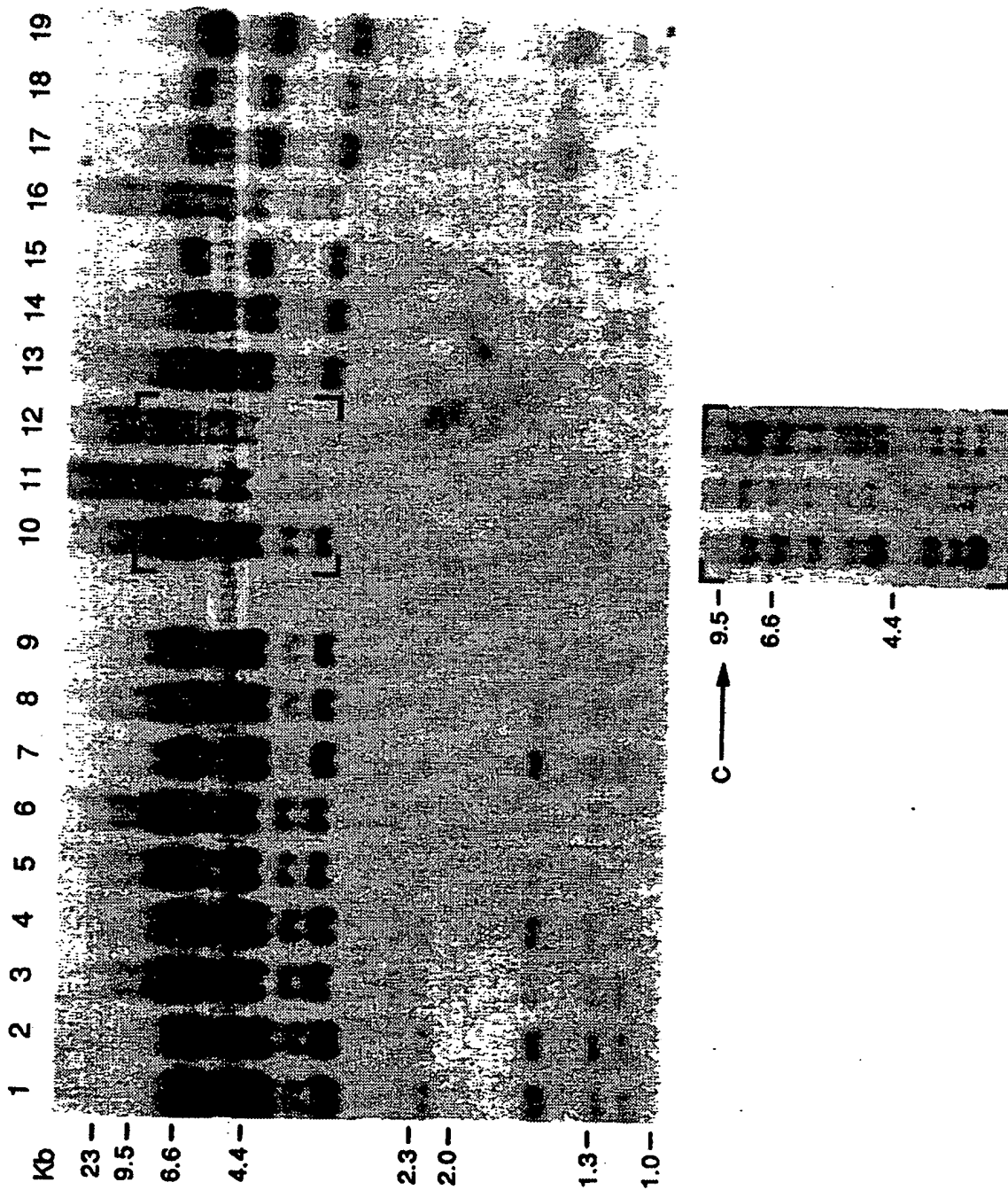


FIG. 6

FIG. 7

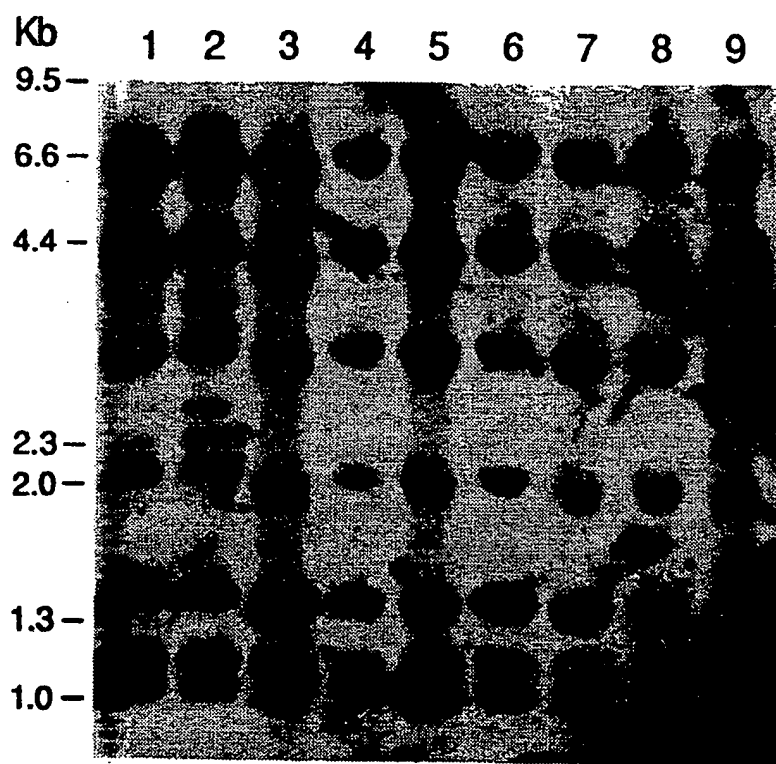
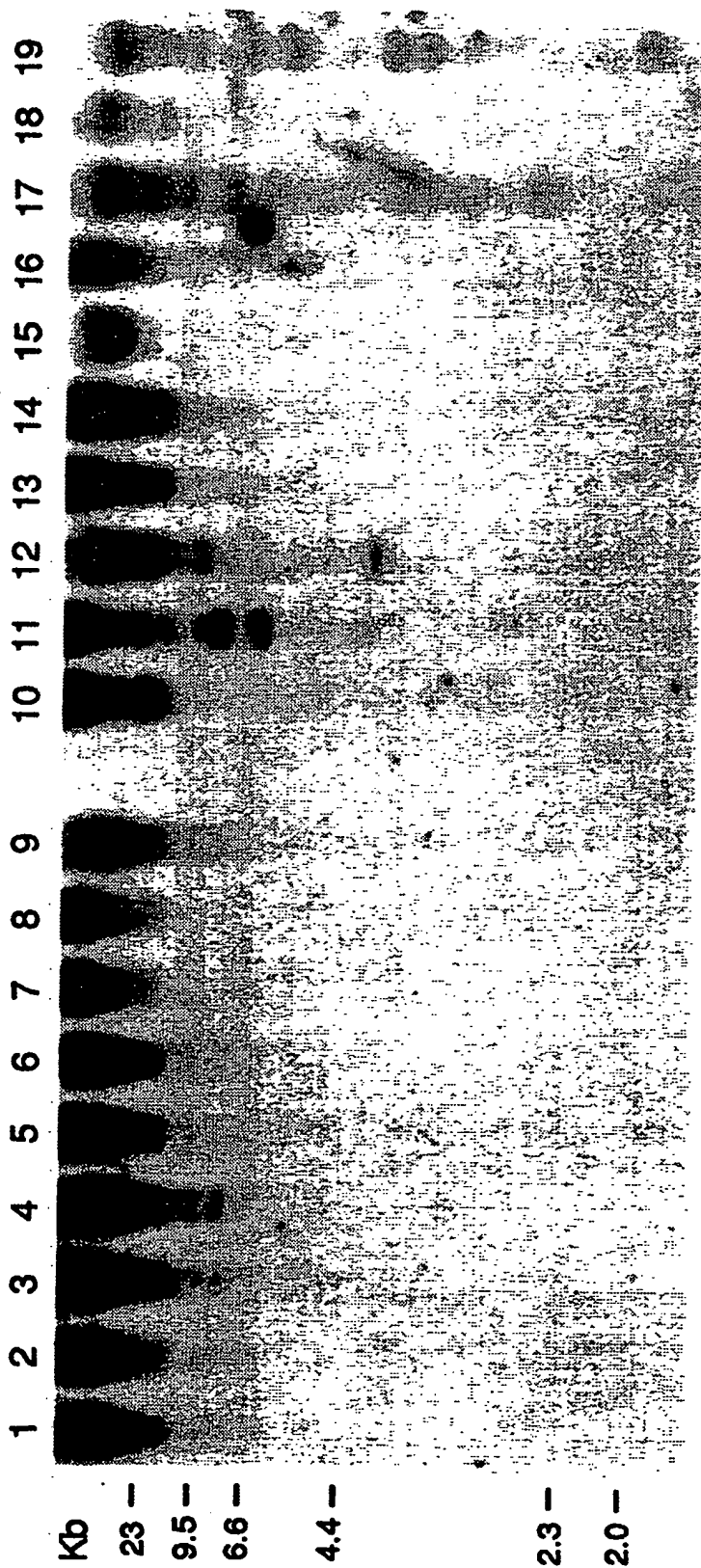


FIG. 8



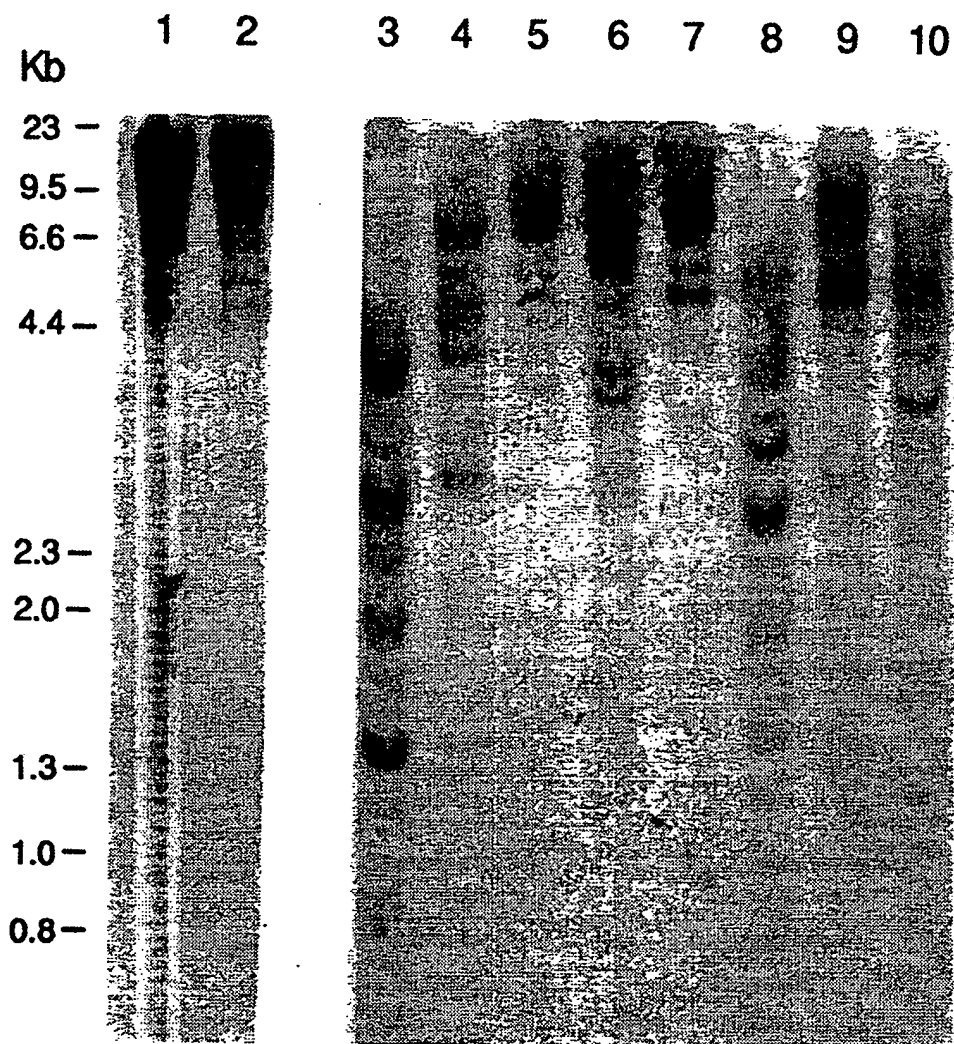


FIG. 9

10 / 18

FIG. 10A:

1	CTT GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG TGT CTG Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Cys Leu
	GCT GGC CGT AGG Ala Gly Arg Arg
61	AGA AGG AGT GTT CAG TGG TGC GCC GTA TCC CAA CCC GAG GCC ACA AAA Arg <u>Arg</u> Ser Val Gln Trp Cys Ala Val Ser Gln Pro Glu Ala Thr Lys Asn
	TGC TTC CAA TGG Cys Phe Gln Trp
121	CAA AGG AAT ATG AGA AAA GTG CGT GGC CCT CCT GTC AGC TGC ATA AAG Gln Arg Asn Met Arg Lys Val Arg Gly Pro Pro Val Ser Cys Ile Lys Leu
	AGA GAC TCC CCC Arg Asp Ser Pro
181	ATC CAG TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC Ile Gln Cys Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr
	CTT GAT GGT GGT Leu Asp Gly Gly



11 / 18

## FIG. 10B

241 TTC ATA TAC GAG GCA GGC CTG GCC CCC TAC AAA CTG CGA CCT GTA GCG  
 Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro Val Ala  
 GCG GAA GTC TAC  
 Ala Glu Val Tyr

301 GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT TAT GCC GTG GCT GTG GTG  
 Gly Thr Glu Arg Gln Pro Arg Thr His Tyr Tyr Ala Val Val Val  
 AAG AAG GGC GGC  
 Lys Lys Gly Gly

361 AGC TTT CAG CTG AAC GAA CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC  
 Ser Phe Gln Leu Asn Glu Leu Gln Gly Leu Lys Ser Cys His Thr Gly  
 CTT CGC AGG ACC  
 Leu Arg Arg Thr

421 GCT GGA TGG AAT GTC CCT ATA GGG ACA CTT CGT CCA TTC TTG AAT TGG  
 Ala Gly Trp Asn Val Pro <sup>C</sup> Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp  
 Thr  
 ACG GGT CCA CCT  
 Thr Gly Pro Pro

12 / 18

## FIG. 10C

481 GAG CCC ATT GAG GCA GCT GTG GCC AGG TTC TTC TCA GCC AGC TGT GTT  
 Glu Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser Cys Val  
 CCC GGT GCA GAT  
 Pro Gly Ala Asp

541 AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG TGT GCG GGG ACA GGG GAA  
 Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys Ala Gly Thr Gly Glu  
 AAC AAA TGT GCC  
 Asn Lys Cys Ala

601 TTC TCC TCC CAG GAA CCG TAC TTC AGC TAC TCT GGT GCC TTC AAG TGT  
 Phe Ser Ser Gln Glu Pro Tyr Phe Ser Tyr Ser Gly Ala Phe Lys Cys  
 CTG AGA GAC GGG  
 Leu Arg Asp Gly  
 Lys

661 GCT GGA GAC GTG GCT TTT ATC AGA GAG AGC ACA GTG TTT GAG GAC CTG  
 Ala Gly Asp Val Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu  
 TCA GAC GAG GCT  
 Ser Asp Glu Ala

13 / 18

## FIG. 10D

721	GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG AAG CCA Glu Arg Asp Glu Tyr Glu Leu Cys Pro Asp Asn Thr Arg Lys Pro
	GTG GAC AAG TTC Val Asp Lys Phe
781	AAA GAC TGC CAT CTG GCC CGG GTC CCT TCT CAT GCC GTT GTG GCA CGA Lys Asp Cys His Leu Ala Arg Val Pro Ser His Ala Val Ala Arg
	AGT GTG AAT GGC Ser Val Asn Gly
841	AAG GAG GAT GCC ATC TGG AAT CTT CTC CGC CAG GCA GAA AAG TTT Lys Glu Asp Ala Ile Trp Asn Leu Leu Arg Gln Ala Gln Glu Lys Phe
	GGA AAG GAC AAG Gly Lys Asp Lys
901	TCA CCG AAA TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG Ser Pro Lys Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu
	CTG TTC AAG GAC Leu Phe Lys Asp

14 / 18

## FIG. 10E

961 TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT GGG CTG  
 Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly Leu  
 TAC CTT GGC TCC  
 Tyr Leu Gly Ser

1021 GGC TAC TTC ACT GCC ATC CAG AAC TTG AGG AAA AGT GAG GAG GAA GTG  
 Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys Ser Glu Glu Val  
 GCT GCC CGG CGT  
 Ala Ala Arg Arg

1081 GCG CGG GTC GTG TGG TGT GCG GTG GGC GAG CAG GAG CTG CGC AAG TGT  
 Ala Arg Val Val Trp Cys Ala Val Gly Glu Gln Glu Leu Arg Lys Cys  
 AAC CAG TGG AGT  
 Asn Gln Trp Ser

1141 GGC TTG AGC GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG  
 Gly Leu Ser Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu  
 TAC CTT GGC TCC  
 Tyr Leu Gly Ser

SUBSTITUTE SHEET

15 / 18

## FIG. 10F

1201 CTG GTG CTG AAA GGA GAA GCT GAT GCC ATG AGT TTG GAT GGA GGA TAT  
 Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Tyr

GTG TAC ACT GCA  
 Val Tyr Thr Ala

T

1261 GGC AAA TGT GGT TTG GTG CCT GTC CTG GCA GAG AAC TAC AAA TCC CAA  
 Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn Tyr Lys Ser Gln  
 Cys

CAA AGC AGT GAC  
 Gln Ser Ser Asp

1321 CCT GAT CCT AAC TGT GTG GAT AGA CCT GTG GAA GGA TAT CTT GCT GTG  
 Pro Asp Pro Asn Cys Val Asp Arg Pro Val Glu Gly Tyr Leu Ala Val  
 GCG GTG GTT AGG  
 Ala Val Val Arg

1381 AGA TCA GAC ACT AGC CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC  
 Arg Ser Asp Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser  
 TGC CAC ACC GCC  
 Cys His Thr Ala

16 / 18

## FIG. 10G

1441	GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG GGC CTG CTC TTC AAC Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Phe Asn
	GC Ala
	CAG ACG GGC TCC Gln Thr Gly Ser
1501	TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC TGT GCC CCT GGG TCT GAC Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys Ala Pro Gly Ser Asp
	CCG AGA TCT AAT Pro Arg Ser Asn
1561	CTC TGT GCT CTG TGT ATT GGC GAC GAG CAG GGT GAG AAT AAG TGC GTG Leu Cys Ala Leu Cys Ile Gly Asp Glu Gln Gly Glu Asn Lys Cys Val
	T CCC AAC AGC AAC Pro Asn Ser Asn
1621	GAG AGA TAC TAC GGC TAC ACT ACT GGG GCT TTC CGG TGC CTG GCT GAG AAT Glu Arg Tyr Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn
	GCT GGA GAC GTT Ala Gly Asp Val

17 / 18

## FIG. 10H

1681 GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT AAC  
Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn Asn

AAT GAG GCA TGG  
Asn Glu Ala Trp

1741 GCT AAG GAT TTG AAG CTG GCA GAC TTT GCG CTG CTG TGC CTC GAT GGC  
Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu Leu Cys Leu Asp Gly

AAA CGG AAG CCT  
Lys Arg Lys Pro

1801 GTG ACT GAG GCT AGA AGC TGC CAT CTT GCC ATG GCC CCG AAT CAT GCC  
Val Thr Glu Ala Arg Ser Cys His Leu Ala Met Ala Pro Asn His Ala

GTG GTG TCT CGG  
Val Val Ser Arg

1861 ATG GAT AAG GTG GAA CGC CTG AAA CAG GTG TTG CTC CAC CAA CAG GCT  
Met Asp Lys Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Ala

AAA TTT GGG AGA  
Lys Phe Gly Arg

SUBSTITUTE SHEET

18 / 18

## FIG. 10I

1921 AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA ACC  
 Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu Thr  
 AAA AAC CTT CTG  
 Lys Asn Leu Leu

1981 TTC AAT GAC AAC ACT GAG TGT CTG GCC AGA CTC CAT GGC AAA ACA ACA  
 Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu His Gly Lys Thr Thr  
 TAT GAA AAA TAT  
 Tyr Glu Lys Tyr

2041 TTG GGA CCA CAG TAT GTC GCA GGC ATT ACT AAT CTG AAA AAG TGC TCA  
 Leu Gly Pro Gln Tyr Val Ala Gly Ile Thr Asn Leu Lys Lys Cys Ser  
 ACC TCC CCC TCC  
 Thr Ser Pro Ser  
 C  
 Leu

2101 TGG AAG CCT GTG AAT TC 2117  
 Trp Lys Pro Val Asn  
 Leu Glu Ala Cys Glu Phe



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US92/04012

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12N 15/00, 15/10, 15/12; A61K 35/20  
US CL : 435/6, 69.1, 320.1; 514/6; 530/395, 400; 536/27

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 69.1, 69.6, 320.1; 514/6; 530/350, 395, 400; 536/27

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, World Patents Index  
search terms: lactoferrin, gene, DNA, cDNA, breast cancer, cancer

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	Clinica Chimica Acta, Vol. 151, issued 1985, W.R. Bezwoda et al, "Enzyme linked immunosorbent assay for lactoferrin. Plasma and tissue measurements", pages 61-69, entire document.	2,9 1,3-8,10-11
X Y	Clinica Chimica Acta, Vol. 157, issued 1986, W.R. Bezwoda et al, "Isolation and characterisation of lactoferrin separate from human whey by adsorption chromatography using Cibacron Blue F3G-A linked affinity adsorbent", pages 89-94, entire document.	2,9 1,3-8
X Y	FEBS Letters, Vol. 109, no. 2, issued January 1980, L. Blackberg et al, "Isolation of lactoferrin from human whey by a single chromatographic step", pages 180-184, entire document.	2,9 1,3-8
Y	J. Sambrook et al., "Molecular cloning techniques, a laboratory manual", published 1989 by Cold Spring Harbor Laboratory Press, pages 12.2-12.15, entire document.	1,3-8

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*A*	Special categories of cited documents: document defining the general state of the art which is not considered to be part of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search

23 July 1992

Date of mailing of the international search report

31 JUL 1992

Name and mailing address of the ISA/  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Authorized officer

DIAN COOK

Facsimile No. NOT APPLICABLE

Telephone No. (703) 308-0196

**INTERNATIONAL SEARCH REPORT**International application No.  
PCT/US92/04012**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Cancer Research, Vol. 46, no. 3, issued March 1986, K. Shirasuna et al, "Isolation and characterization of different clones including myoepithelial-like variants from a clonal neoplastic epithelial duct cell line of human salivary gland origin", pages 1418-1426, especially abstract.	12-14